

doi:10.3969/j.issn.1005-3697.2026.04.003

✧智能医学研究专题✧

# 基于双期增强 CT 的多模态模型早期预测急性胰腺炎相关急性肺损伤

陈姝君,熊园,宋学亮,邓萍,邓鸿,张小明,李兴辉

(川北医学院附属医院放射科·四川省医学影像重点实验室,四川南充 637000)

**【摘要】目的:**构建基于腹部双期增强计算机断层扫描(CECT)的多模态深度学习-影像组学-临床(DRC)模型以早期预测急性胰腺炎相关急性肺损伤/急性呼吸窘迫综合征(AP-ALI/ARDS)。**方法:**回顾性纳入 320 例 AP 患者为研究对象,按 7:3 比例随机划分为训练集( $n=189$ )和内部测试集( $n=81$ ),并设外部测试集( $n=50$ )。收集患者入院 48 h 内临床资料及 CECT 图像,采用逻辑回归方法筛选临床独立危险因素。手动勾画动脉期及门静脉期胰腺实质并提取筛选影像组学与深度学习特征。采用随机森林算法分别构建临床模型、影像组学模型、深度学习模型及 DRC 模型。采用受试者工作特征(ROC)曲线、DeLong 检验、校准曲线与决策曲线分析(DCA)比较模型性能。**结果:**320 例 AP 患者中,109 例(34.06%)发生 AP-ALI/ARDS。AP 严重程度和血糖水平为 AP-ALI/ARDS 相关独立危险因素。最终筛选出 6 个影像组学特征和 19 个深度学习特征。DRC 模型展现了卓越的预测性能,在训练集、内部及外部测试集中 ROC 曲线下面积(AUC)分别为 0.973、0.934、0.920。DeLong 检验表明 DRC 模型的 AUC 显著高于临床模型、影像组学模型( $P<0.05$ )。校准曲线表明 DRC 模型具有最佳性能,且 DCA 显示 DRC 模型在大多数阈值概率下相比其他模型能提供更高的总体净收益。**结论:**DRC 模型在早期预测 AP-ALI/ARDS 方面具有优异性能,可为临床干预提供决策支持。

**【关键词】**急性胰腺炎;急性肺损伤;深度学习;急性呼吸窘迫综合征;影像组学

**【中图分类号】**R445.3 **【文献标志码】**A

## Early prediction of acute pancreatitis-induced acute lung injury using a multimodal model based on dual-phase contrast-enhanced CT

CHEN Shu-jun, XIONG Yuan, SONG Xue-liang, DENG Ping, DENG Hong, ZHANG Xiao-ming, LI Xing-hui

(Medical Imaging Key Laboratory of Sichuan Province, Department of Radiology, Affiliated Hospital of North Sichuan Medical College, Nanchong 637000, Sichuan, China)

**【Abstract】Objective:** To develop a multimodal deep learning-radiomics-clinical (DRC) model based on abdominal dual-phase contrast-enhanced CT (CECT) for the early prediction of acute pancreatitis-associated acute lung injury/acute respiratory distress syndrome (AP-ALI/ARDS). **Methods:** 320 AP patients were enrolled and randomly divided into a training set ( $n=189$ ) and an internal test set ( $n=81$ ) in a 7:3 ratio, with an additional external test set ( $n=50$ ). Clinical data and CECT images acquired within 48 hours after admission were collected. Independent clinical risk factors were selected using Logistic regression. The pancreatic parenchyma was manually delineated on both arterial and portal venous phase images, from which radiomics and deep learning features were subsequently extracted and screened. The random forest algorithm was used to construct the clinical model, radiomics model, deep learning model, and DRC model, respectively. Model performance was evaluated and compared using receiver operating characteristic (ROC) curves, the DeLong test, calibration curves, and decision curve analysis (DCA). **Results:** Among the 320 AP patients, 109 (34.06%) developed AP-ALI/ARDS. AP severity and blood glucose level were identified as independent risk factors. 6 radiomic features and 19 deep learning features were ultimately selected. The DRC model demonstrated superior predictive performance, achieving areas under the ROC curve (AUCs) of 0.973, 0.934, and 0.920 in the training, internal test, and external test sets, respectively. The DeLong test indicated that the AUC of the DRC model was significantly higher than those of the clinical and radiomics models ( $P<0.05$ ). Calibration curves demon-

**基金项目:**国家自然科学基金(82371961);四川省卫生健康委员会医学科技项目(24QNMP062);川北医学院附属医院科研项目(BS20211116)

**作者简介:**陈姝君(2000-),女,硕士研究生。E-mail:junoehen0426@163.com

**通讯作者:**李兴辉,博士。E-mail:lixinghui1005@126.com

strated that the DRC model exhibited the best performance, and DCA showed that it provided a greater overall net benefit across most threshold probabilities compared to other models. **Conclusion:** The DRC model exhibits excellent performance for the early prediction of AP-ALI/ARDS and shows potential as decision-support tool to guide clinical intervention.

**【Key words】** Acute pancreatitis; Acute lung injury; Deep learning; Acute respiratory distress syndrome; Radiomics

急性胰腺炎(acute pancreatitis, AP)是由胰腺内消化酶失调引发的急性炎症性疾病,可导致胰腺组织水肿、坏死,甚至引发胰腺外器官的炎症和损伤,大约 20% 的 AP 患者进展为重度 AP (severe acute pancreatitis, SAP)<sup>[1]</sup>。AP 引发的急性肺损伤或急性呼吸窘迫综合征(acute pancreatitis-induced acute lung injury/acute respiratory distress syndrome, AP-ALI/ARDS)是 SAP 患者最常见和最早的器官功能损害,发生率为 15%~60%,死亡率高达 60%<sup>[2-3]</sup>。

AP 引发炎症级联反应促使炎症介质和炎症细胞在肺内蓄积,导致肺部正常肺泡-毛细血管屏障功能受损,引起肺水肿和弥漫性肺泡损伤,严重损害肺内的血气交换。AP-ALI/ARDS 早期症状隐匿,患者可能在 AP 发病后数小时至数天内出现呼吸功能障碍,甚至可能进展为呼吸衰竭<sup>[4-5]</sup>。目前,AP-ALI/ARDS 的诊断主要依据临床症状、血氧饱和度监测和动脉血气分析。治疗方案仍局限于液体管理、机械通气等支持措施,这一现状使得早期识别与干预变得至关重要<sup>[6]</sup>。既往 AP-ALI/ARDS 相关的研究取得一定进展,年龄和吸烟可作为 AP 相关呼吸衰竭的预测因子<sup>[7]</sup>。另有研究构建反向传播人工神经网络模型预测评估 SAP-ARDS 风险及严重程度<sup>[8]</sup>。Jia 等<sup>[9]</sup>基于常规临床特征和 SAP 血浆生物标志物建立预测模型。亦有学者<sup>[10]</sup>构建简化肺损伤相关评分以识别持续性急性呼吸衰竭的风险。然而,很少有研究纳入所有严重程度 AP 人群同时使用易获得的参数以预测 AP-ALI/ARDS。近年来,医学影像与人工智能相结合可早期预测 AP 的胰腺坏死、严重程度、相关器官损伤(肾脏损伤、肝脏损伤),在协助 AP 患者管理决策方面显示出巨大潜力<sup>[11-14]</sup>。在 AP-ALI/ARDS 发生预测方面,尚无融合影像数据且适用于所有严重程度 AP 患者的有效预测工具。

本研究旨在基于腹部双期 CECT 融合深度学习、影像组学及临床特征,构建可在 AP 早期阶段预测 AP-ALI/ARDS 发生的多模态模型,以期通过早期风险识别为启动针对性治疗提供决策支持。

## 1 资料与方法

### 1.1 一般资料

回顾性纳入 2020 年 1 月至 2024 年 12 月川北

医学院附属医院(中心 1)和自贡市第四人民医院(中心 2)确诊 AP 的住院患者为研究对象。本研究经中心 1 和中心 2 伦理委员会批准免除知情同意,批准号为[2021]08。

纳入标准:(1)符合 2012 年修订的亚特兰大分类(the revised atlanta classification, RAC)AP 的诊断要求<sup>[15]</sup>;(2)年龄 $\geq 18$  岁;(3)症状发作至入院时间 $\leq 48$  h;(4)入院 48 h 内完成腹部双期 CECT 检查;(5)临床资料、实验室数据以及图像信息完整;(6)发病后 7 d 内接受动脉血气分析检查。排除标准:(1)心源性肺水肿、心力衰竭;(2)转院患者;(3)慢性阻塞性肺疾病急性加重,哮喘以及活动性肺结核等呼吸系统疾病;(4)患有感染性疾病、传染性疾病、恶性肿瘤、严重慢性消耗性疾病;(5)在 1 年内接受过腹部、盆腔手术治疗;(6)CECT 影像质量差;(7)既往胰腺手术病史;(8)慢性胰腺炎急性发作或复发性急性胰腺炎;(9)妊娠期或哺乳期妇女。

本研究共纳入 320 例 AP 患者,按照 7:3 的比例将中心 1 患者( $n=270$ )随机分为训练集( $n=189$ )和内部测试集( $n=81$ ),中心 2 患者( $n=50$ )作为外部测试集。

### 1.2 AP-ALI/ARDS 诊断标准

本研究关于 AP-ALI/ARDS 的诊断需满足 2012 年柏林定义的所有标准<sup>[16]</sup>:(1)起病后 7 d 内急性发作;(2)胸部影像学(X 射线或 CT)表现的双侧混浊不能完全解释为积液、肺叶/肺塌陷或结节;(3)出现氧合障碍,动脉血氧分压( $\text{PaO}_2$ )/吸入氧浓度( $\text{FiO}_2$ ) $\leq 300$  mmHg,呼气末正压通气或持续气道正压通气 $\geq 5$  cmH<sub>2</sub>O。患有心源性肺水肿、既往存在间质性肺病或其他可识别的 ALI/ARDS 原因的患者被排除在外。基于上述标准,入组患者分为 AP-ALI/ARDS 和非 AP-ALI/ARDS。

### 1.3 临床和图像数据收集

从电子病历系统收集患者入院 48 h 内的临床数据,基本信息包括年龄、性别、病因、RAC 严重程度、改良 CT 严重程度指数(modified CT severity index, mCTSI)评分<sup>[17]</sup>、AP 严重性床边指数(side index for severity in acute pancreatitis, BISAP)<sup>[18]</sup>等;实验室指标包括血常规、胰腺炎相关生化指标等。

从医学影像存储与传输系统(picture archiving and communication system, PACS)导出腹部双期

CECT 图像,所有患者均采用以下任一扫描仪进行腹部双期增强 CT 扫描:Somatom Force(美国西门子)、Somatom Definition AS+ 128(美国西门子)或 Light Speed VCT 128(美国 GE)。患者取仰卧位,头先进,双臂上举。扫描范围自膈顶至髂嵴水平,扫描参数:管电压 120 kV,管电流 150~250 mAs,矩阵  $512 \times 512$ ,视野  $350 \text{ mm} \times 350 \text{ mm}$ ,螺距 0.6~1.0,转速 0.25~0.5 s/r,层厚 5 mm,层间距 5 mm。平扫结束后,经肘正中静脉留置针使用高压注射器(Ulrich CT Plus 150, Ulrich Medical)注射非离子型碘对比剂,剂量为 1.5 mL/kg,注射速率 2.5~3.5 mL/s,随后以 30~40 mL 生理盐水冲管。动脉期及门静脉期图像分别在触发后延迟 20~30 s、50~60 s 采集。基于上述基线数据,采用单因素及多因素 Logistic 回归分析确定 AP-ALI/ARDS 的临床独立危险因素。

#### 1.4 特征提取与筛选

本研究有关影像组学特征的提取流程均符合图像生物标志物标准化倡议(image biomarker standardization initiative, IBSI)<sup>[19-20]</sup>。将 DICOM 格式的层厚 5 mm 的图像导入 3D Slicer 软件(V 5.0.3, <https://www.slicer.org/>),对临床分组信息设盲,由具有 6 年腹部 CECT 影像诊断经验的放射科医师 1 在所有图像上逐层手动勾画胰腺实质体积感兴趣区(volume of interest, VOI),避开胰管、周围血管、肠管及脂肪组织。为减少多中心、多扫描设备带来的中心效应并提升模型泛化能力,本研究对图像进行了重插值、灰度离散化和滤波器等标准化预处理操作。使用 Pyradiomics 工具(<http://www.radiomics.io/pyradiomics.html> 3.0.1 版)分别从动脉期和门静脉期的胰腺实质 VOI 中提取 7 类影像组学特征组:一阶(first order)、灰度共生矩阵(gray level co-occurrence matrix, GLCM)、灰度依赖矩阵(gray level dependence matrix, GLDM)、灰度游程长度矩阵(gray level run length matrix, GLRLM)、灰度大小区域矩阵(gray level size zone matrix, GLSZM)、邻近灰度色调差异矩阵(neighbouring gray tone difference matrix, NGTDM)、形状(shape)。为评估影像组学特征的可重复性,本研究选取 50 例患者图像进行可靠性分析:由医师 1 间隔 1 个月进行第二次勾画(组内),并由具有 8 年腹部 CECT 影像诊断经验的放射科医师 2 进行独立勾画(组间)。计算并保留组内和组间相关系数(inter-class and intraclass correlation coefficients, ICCs)均 > 0.75 的具有较高一致性的特征<sup>[21]</sup>。

基于 uAI Research Portal(联影,中国,

V 20250430),利用平台内置的 18 层深度 3D Residual Net-18(3D ResNet-18)作为骨干网络进行深度学习训练并提取深度学习特征。为进一步提升模型性能并抑制过拟合,在训练过程中引入了数据增强技术,包括随机翻转、亮度调整和对比度增强,同时采用均衡采样以保持类别平衡。在最后一次卷积运算后,通过输出层的全局平均池化层将特征图转换为深度学习一维向量特征。

对所有特征进行 Z-score 标准化后,采用分步策略进行特征筛选:对 ICCs > 0.75 的影像组学特征,通过独立样本 t 检验或 Mann-Whitney U 检验筛选出  $P < 0.05$  的最佳影像组学特征;对于深度学习特征采用最小冗余最大相关(minimum redundancy maximum relevance, mRMR)算法,剔除相关性高于 0.9 的冗余特征后,保留重要性排名前 20 的特征。最后,应用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)解决多重共线性问题并筛选出最佳特征集。所有筛选步骤均在训练集内完成,以避免数据泄露。

#### 1.5 模型的开发与评估

基于最佳影像组学特征及最佳深度学习特征,本研究分别计算影像组学评分(radiomics\_score, Rad\_score)与深度学习评分(deep learning\_score, DL\_score)。独立构建临床、影像组学及深度学习三个单模态模型后,采用特征层融合策略整合多模态信息,构建最终的多模态深度学习-影像组学-临床(deep learning-radiomics-clinical, DRC)联合模型。采用随机森林(random forest, RF)机器学习分类算法测试模型性能。为评估比较模型性能与可靠性,本研究计算受试者工作特征(receiver operating characteristic, ROC)曲线下面积(areas under the ROC curve, AUC)、准确度、敏感度、特异度、精确度及 F1-score。采用 DeLong 检验比较模型间 AUC 差异,同时通过校准曲线与 Brier score 评估预测概率的一致性,并利用决策曲线分析(decision curve analysis, DCA)考察模型的临床净收益<sup>[22]</sup>。最后,基于 SHAP 可解释性工具对最优模型进行可视化分析以提供直观的预测工具。上述流程均使用 Python 3.12.0 完成。

#### 1.6 统计学分析

采用 SPSS 26.0 (IBM Corporation, Armonk, USA)分析 AP-ALI/ARDS 和非 AP-ALI/ARDS 患者临床数据的组间差异。连续变量经正态性检验后,符合正态分布者以  $(\bar{x} \pm s)$  描述,组间比较采用独立样本 t 检验;非正态分布者以  $[M(Q_1, Q_3)]$  描述,组间比较采用 Mann-Whitney U 检验。分类变

量以 $[n(\%)]$ 描述,组间比较采用独立样本 $\chi^2$ 检验。 $P<0.05$ 为差异有统计学意义。

## 2 结果

### 2.1 一般数据比较

本研究选取了中心 1( $n=270$ )和来自中心 2( $n=50$ )的 320 例 AP 患者的临床数据,轻症、中度重症及重症分别占比 34.69%、48.75%、16.56%。AP-ALI/ARDS 共有 109 例(34.06%),非 AP-ALI/ARDS 共有 211 例(65.94%)。在训练集、两个测试集中,RAC 严重程度、白细胞、中性粒细胞、淋巴细胞、血糖、mCTSI、BISAP 均有统计学差异( $P<0.05$ )。见表 1。

### 2.2 特征筛选和模型构建

基于训练集临床数据的单因素及多因素 Logistic 回归分析确定,RAC 严重程度与血糖是 AP-ALI/ARDS 相关的临床独立危险因素以构建临床模型(表 2)。从动脉期和门静脉期胰腺实质 VOI 中分别提取 1 170 个影像组学特征,最终筛选出 6 个系数非零的最佳影像组学特征,计算 Rad\_Score 用于构建影像组学模型(表 3)。利用全局平均池化层提取并筛选保留 19 个最佳深度学习特征,计算 DL\_score 用于构建深度学习模型(表 4)。最后,通过特征层融合临床独立危险因素、影像组学评分及深度学习评分开发多模态 DRC 模型。基于训练组数据结合 10 折交叉验证调整优化,最终确定四个模型中 RF 算法的最佳参数设置。

表 1 训练集和测试集的一般资料比较 $[M(Q_1, Q_3), n(\%)]$

资料	训练集 ( $n=189$ )		$\chi^2/Z$ 值	$P$ 值	内部测试集 ( $n=81$ )		$\chi^2/Z$ 值	$P$ 值	外部测试集 ( $n=50$ )		$\chi^2/Z$ 值	$P$ 值
	非 AP-ALI/ARDS ( $n=125$ )	AP-ALI/ARDS ( $n=64$ )			非 AP-ALI/ARDS ( $n=54$ )	AP-ALI/ARDS ( $n=27$ )			非 AP-ALI/ARDS ( $n=32$ )	AP-ALI/ARDS ( $n=18$ )		
性别			1.753	0.185			0.437	0.509			1.719	0.190
男	77 (61.60)	33 (51.56)			34 (62.96)	19 (70.37)			22 (68.75)	9 (50.00)		
女	48 (38.40)	31 (48.44)			20 (37.04)	8 (29.63)			10 (31.25)	9 (50.00)		
年龄(岁)	47.00 (38.00,58.00)	54.00 (35.00,68.75)	-1.100	0.271	49.00 (37.75,59.25)	51.00 (43.00,72.00)	-1.558	0.119	43.00 (36.25,50.00)	47.50 (38.50,62.25)	-1.072	0.284
病因			7.558	0.056			4.229	0.238			1.423	0.700
酒精性	17 (13.60)	1 (1.56)			14 (25.93)	2 (7.41)			4 (12.50)	3 (16.67)		
胆源性	29 (23.20)	19 (29.69)			11 (20.37)	8 (29.63)			9 (28.12)	6 (33.33)		
高甘油三酯血症性	50 (40.00)	26 (40.63)			17 (31.48)	11 (40.74)			11 (34.38)	7 (38.89)		
其他	29 (23.20)	18 (28.12)			12 (22.22)	6 (22.22)			8 (25.00)	2 (11.11)		
RAC 严重程度			33.735	<0.001			18.573	<0.001			17.737	<0.001
轻症	58 (46.40)	4 (6.25)			30 (55.56)	3 (11.11)			16 (50.00)	0 (0)		
中度重症	53 (42.40)	40 (62.50)			22 (40.74)	17 (62.96)			14 (43.75)	10 (55.56)		
重症	14 (11.20)	20 (31.25)			2 (3.70)	7 (25.93)			2 (6.25)	8 (44.44)		
胰淀粉酶(U/L)	216.00 (76.50,964.50)	398.00 (134.75,1247.00)	-1.377	0.169	157.00 (64.50,469.75)	264.00 (122.00,792.00)	-1.894	0.058	238.00 (91.50,595.00)	633.00 (193.00,1468.50)	-2.375	0.018
脂肪酶(U/L)	331.0 (135.00,1396.50)	541.50 (190.50,1394.75)	-1.115	0.265	372.50 (110.75,829.25)	402.00 (172.00,1397.00)	-1.348	0.178	487.50 (161.00,1221.50)	867.00 (356.50,2128.00)	-1.617	0.106
血清淀粉酶(U/L)	264.00 (91.00,790.75)	410.00 (145.00,1339.25)	-1.280	0.201	226.00 (97.25,576.50)	597.00 (141.00,1001.00)	-2.034	0.042	344.00 (103.25,790.75)	703.00 (173.25,1154.25)	-1.910	0.056
白细胞( $\times 10^9/L$ )	13.13 (9.96,15.96)	14.33 (11.43,18.18)	-2.460	0.014	13.27 (10.91,15.84)	16.22 (13.05,18.42)	-2.249	0.024	13.38 (10.09,15.88)	16.31 (12.94,18.58)	-2.011	0.044
中性粒细胞( $\times 10^9/L$ )	10.64 (8.07,14.59)	13.03 (10.16,16.74)	-3.122	0.002	11.58 (8.77,14.12)	14.42 (12.04,16.09)	-2.650	0.008	10.53 (8.42,11.75)	12.49 (10.42,15.35)	-2.516	0.012
淋巴细胞( $\times 10^9/L$ )	1.22 (0.74,1.73)	0.87 (0.59,1.38)	-2.737	0.006	1.34 (0.95,1.74)	0.96 (0.65,1.45)	-2.350	0.019	1.23 (1.05,1.46)	0.98 (0.84,1.38)	-2.113	0.035
Na <sup>+</sup> (mmol/L)	136.50 (133.80,138.55)	136.80 (132.40,139.00)	-0.451	0.652	136.00 (132.08,139.23)	136.10 (130.80,138.10)	-0.516	0.606	136.20 (134.23,138.60)	136.25 (133.43,137.30)	-1.162	0.245
K <sup>+</sup> (mmol/L)	3.78 (3.52,3.99)	3.77 (3.50,4.27)	-0.420	0.674	3.75 (3.54,4.05)	3.56 (3.30,3.91)	-1.593	0.111	3.80 (3.44,4.15)	3.82 (3.48,4.01)	-0.142	0.887
血糖(mmol/L)	7.07 (6.08,9.14)	10.32 (7.65,13.40)	-5.354	<0.001	7.57 (6.40,10.55)	10.40 (7.61,12.96)	-2.244	0.025	7.64 (7.00,11.31)	11.60 (7.85,18.77)	-2.163	0.034
mCTSI	4.00 (2.00,4.00)	4.00 (4.00,6.00)	-3.648	<0.001	4.00 (2.00,4.50)	4.00 (4.00,6.00)	-2.255	0.024	4.00 (2.00,5.50)	6.00 (4.00,6.00)	-2.417	0.016
BISAP	0.00 (0.00,1.00)	1.00 (0.00,2.00)	-2.935	0.003	1.00 (0.00,1.00)	1.00 (1.00,2.00)	-2.578	0.010	1.00 (0.00,2.00)	2.00 (1.00,2.00)	-2.137	0.033

表 2 单因素/多因素 Logistic 回归分析筛选临床独立危险因素

变量	单因素		多因素	
	OR(95% CI)	P 值	OR(95% CI)	P 值
RAC 严重程度	4.03 (2.38~6.83)	<0.001	2.71 (1.39~5.28)	0.004
白细胞	1.08 (1.01~1.15)	0.014		
中性粒细胞	1.05 (0.99~1.10)	0.002		
淋巴细胞	0.48 (0.29~0.80)	0.006		
血糖	1.32 (1.19~1.46)	<0.001	1.29 (1.14~1.46)	<0.001
mCTSI	1.43 (1.18~1.73)	<0.001		
BISAP	1.37 (0.99~1.10)	0.003		

表 3 最佳影像组学特征及其相关系数

期相来源	特征名称	相关系数
动脉期	log_glcml_log-sigma-4-0-mm-3D-Correlation	-0.008 125 250
	log_glcml_log-sigma-4-0-mm-3D-ld	0.034 546 234
	wavelet_firstorder_wavelet-LHL-Kurtosis	-0.037 081 870
门静脉期	log_firstorder_log-sigma-2-0-mm-3D-10Percentile	0.043 388 540
	log_firstorder_log-sigma-4-0-mm-3D-Kurtosis	0.020 970 725
	wavelet_firstorder_wavelet-LLL-Kurtosis	-0.012 985 542

表 4 最佳深度学习特征及其相关系数

期相来源	特征名称	相关系数
动脉期	A13	-0.163 155 571
	A43	-0.067 461 394
	A44	0.013 573 500
	A62	0.028 027 596
	A70	0.036 438 175
	A102	-0.116 493 000
	A131	-0.013 833 231
	A161	-0.079 426 760
	A184	0.075 524 670
	A187	-0.075 711 930
	A205	0.060 447 767
	A212	-0.244 722 515
	A232	-0.106 526 770
	门静脉期	V35
V46		0.518 681 500
V60		0.078 697 370
V78		0.147 534 445
V91		-0.086 713 105
V110		-0.107 426 600

2.3 不同模型的性能评估与比较

训练集、内部测试集和外部测试集中四个模型的 ROC 曲线见图 1A、图 1D 及图 1G, AUC、敏感度等性能结果如表 5 所示。DRC 模型在三个数据集均显示出最好的 AP-ALI/ARDS 预测性能, AUC 分别为 0.973 (95% CI: 0.954~0.992), 0.934 (95% CI: 0.868~0.999) 和 0.920 (95% CI: 0.848~0.992)。其他评估指标包括准确度、敏感度、特异度、精确度及 F1-score, 一致表明 DRC 模型优于其余三个模型。DeLong 检验证实 DRC 模型与临床模型在内部及外部测试集中的 AUC 差异均有统计学意义 ( $P < 0.05$ ), DRC 模型与影像组学模型在内

部测试集中存在差异 ( $P < 0.05$ )。校准曲线表明 DRC 模型展现出最优的校准性能 (Brier score: 0.076, 0.096, 0.110), 与实际观察结果进行了良好的校准, 见图 1B、图 1E 及图 1H。DCA 进一步表明 DRC 模型与其余模型相比, 在大多数阈值范围内提供了更卓越的临床净收益。见图 1C、图 1F 及图 1I。

基于 DRC 模型构建了可视化 SHAP 图, 可快速、直观地评估各特征重要性分布及范围值, 见图 2。DRC 模型变量重要性分析表明 DL\_score 对预测结果的贡献度最高 (45.2%), 其次为 Rad\_score (31.6%) 与临床独立危险因素 (23.2%)。

表 5 单模态和多模态模型在训练集和内、外部测试集中的性能

模型	训练集						内部测试集						外部测试集								
	AUC (95% CI)	准确率	敏感度	特异度	精确度	F1-score	Brier score	AUC (95% CI)	准确率	敏感度	特异度	精确度	F1-score	Brier score	AUC (95% CI)	准确率	敏感度	特异度	精确度	F1-score	Brier score
临床模型	0.848 (0.794~0.901)	0.781	0.720	0.741	0.588	0.671	0.158	0.821 (0.706~0.935)	0.852	0.630	0.704	0.535	0.657	0.159	0.793 (0.661~0.926)	0.800	0.686	0.720	0.522	0.632	0.180
影像组学模型	0.873 (0.819~0.927)	0.781	0.784	0.783	0.649	0.709	0.121	0.815 (0.711~0.918)	0.741	0.778	0.765	0.625	0.678	0.168	0.825 (0.703~0.948)	0.882	0.788	0.820	0.682	0.769	0.177
深度学习模型	0.909 (0.868~0.951)	0.844	0.832	0.836	0.720	0.777	0.113	0.875 (0.797~0.953)	0.852	0.778	0.803	0.657	0.742	0.132	0.886 (0.796~0.976)	0.867	0.800	0.820	0.650	0.743	0.128
DRC 模型	0.973 (0.954~0.992)	0.922	0.912	0.915	0.843	0.881	0.076	0.934 (0.868~0.999)	0.889	0.870	0.877	0.774	0.828	0.096	0.920 (0.848~0.992)	0.867	0.829	0.840	0.684	0.765	0.110

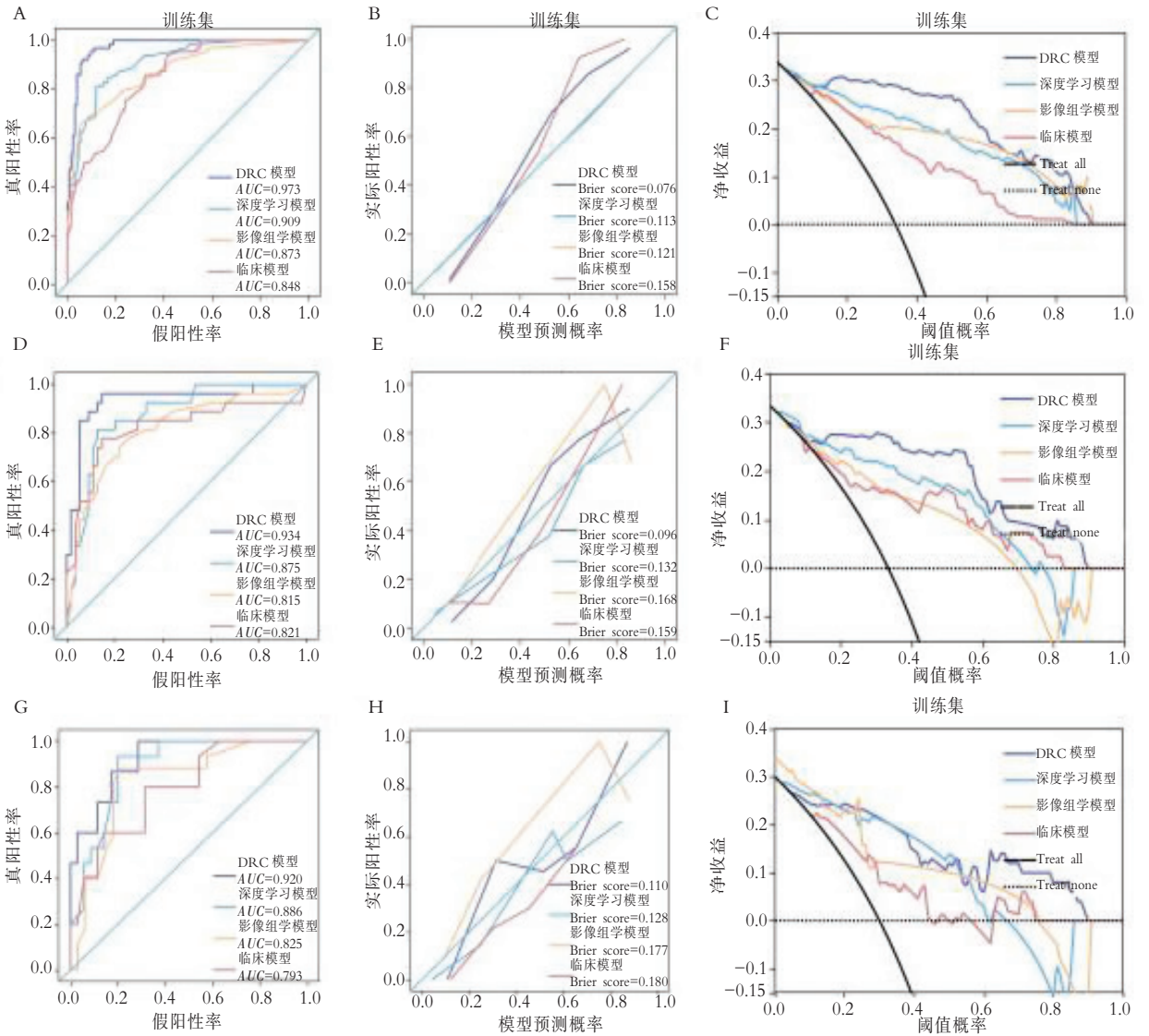


图 1 单模态与多模态模型在训练集和内、外部测试集中的 ROC 曲线、校准曲线和决策曲线

A、D、G 为 ROC 曲线; B、E、H 为校准曲线; C、F、I 为决策曲线。

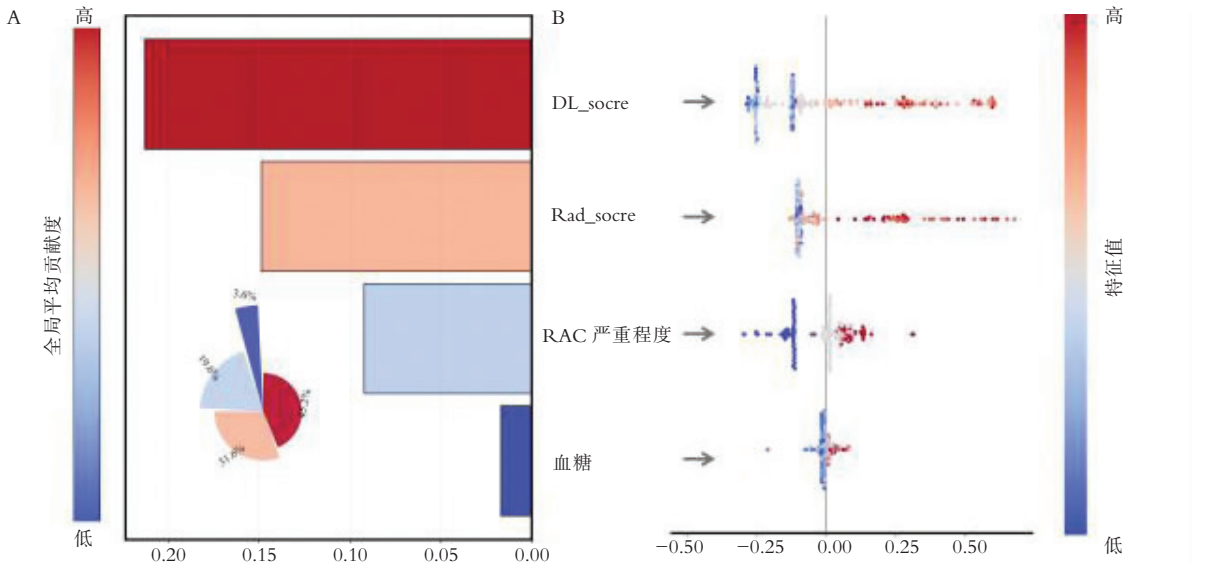


图 2 深度学习-影像组学-临床(DRC)模型的 SHAP 可解释性分析

A. 全局 SHAP 摘要图:平均绝对 SHAP 值量化各特征的全局贡献度; B. SHAP 蜂群图:展示各特征 SHAP 值的分布,颜色表示特征值的高低,红色高值,蓝色低值。

## 2.4 DRC 模型预测结果验证案例

为进一步直观展示多模态 DRC 模型的预测效

能,本研究选取了两例代表性病例进行影像学验证。见图 3 及图 4。

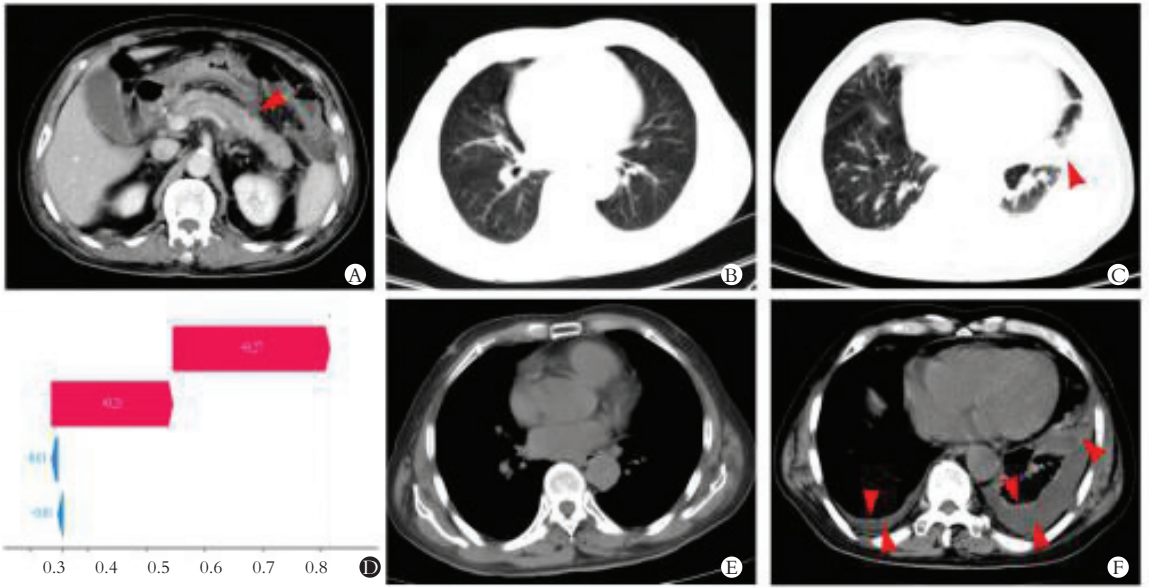


图 3 病例 1(模型正确预警),男性患者,70 岁

发病第 2 天(入院第 1 天)CECT 显示胰腺体积稍肿胀,周围脂肪间隙模糊,可见片絮状渗出影,左侧肾前筋膜增厚(图 3A),同期胸部 CT 未见明确异常征象(图 3B、图 3E)。DRC 模型预测其发生 AP-ALI/ARDS 的风险概率为 0.85(高风险),见图 3D SHAP 瀑布图。随访至发病第 4 天(入院第 3 天),复查胸部 CT 显示双侧胸腔少量积液、胸膜增厚,左肺下叶局部实变不张(图 3C、图 3F)。患者满足 AP-ALI/ARDS 诊断标准,证实模型早期预警的准确性。

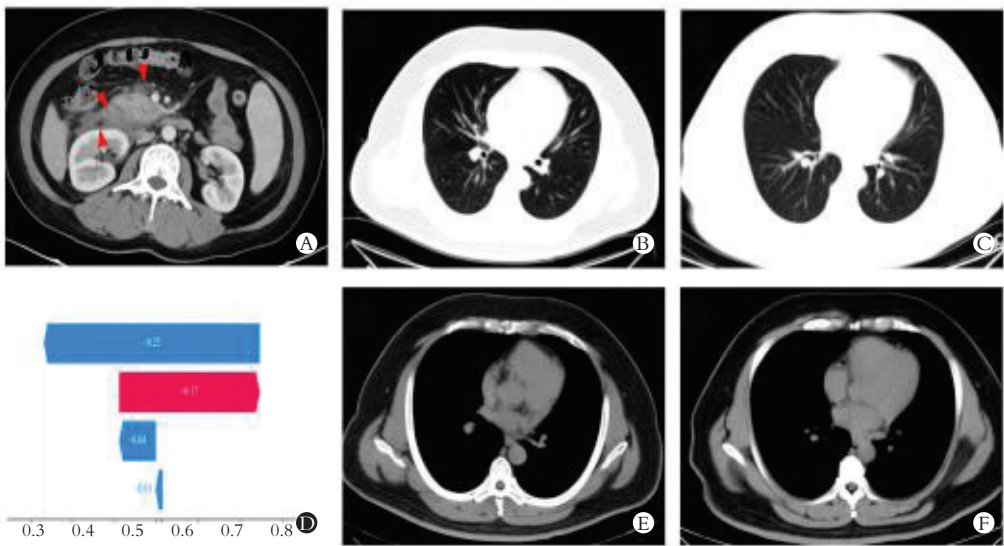


图 4 病例 2(模型正确排除),男性患者,47 岁。

发病第 1 天(入院第 1 天)CECT 显示胰头颈部体积增大,实质密度稍减低,胰腺周围脂肪间隙模糊,呈渗出改变,邻近胰周肠系膜、肾周筋膜增厚、粘连,周围并见絮片状模糊影及索条影(图 4A)。同期胸部 CT 未见异常(图 4B、图 4E)。DRC 模型预测 AP-ALI/ARDS 发生风险概率为 0.32(低风险),见图 4D SHAP 瀑布图。随访至发病第 5 天(入院第 5 天)复查胸部 CT 仍未见明显异常(图 4C、图 4F),与模型预测结果一致。患者不满足 AP-ALI/ARDS 诊断标准,证实模型早期预警的准确性。

## 3 讨论

### 3.1 研究创新性与多模态 DRC 模型的整体效能

本研究首次基于多中心双期 CECT 影像,融合临床、影像组学及深度学习三类特征,构建早期预测 AP-ALI/ARDS 的多模态联合 DRC 模型。DRC 模

型在训练集和内、外部测试集中均表现出强大的预测效能(AUC:0.973,0.934,0.920),其优于单模态的临床、影像组学和深度学习模型,表明深度挖掘医学影像特征并与临床特征相结合,能够实现信息互补与综合性能优化。

### 3.2 聚焦关键临床危险因素独立预测价值及其机制

既往研究<sup>[8-9,23]</sup>均基于 SAP 患者人群临床特征构建 AP-ALI/ARDS 预测模型,未整合与挖掘医学影像数据潜力。本研究涵盖轻症、中度重症及重症的全严重程度 AP 患者,临床模型取得  $AUC = 0.848$  的良好预测性能,更符合真实的临床场景,提升模型的普适性。这也解释了本研究临床模型  $AUC$  低于以往仅基于 SAP 人群研究的原因。本研究发现 AP 严重程度 ( $OR = 2.71$ ) 与血糖水平 ( $OR = 1.29$ ) 为 AP-ALI/ARDS 的独立危险因素,提示 AP 初始严重程度和糖代谢紊乱在 AP-ALI/ARDS 发生中可能具有重要作用。此外,多项研究<sup>[24-25]</sup>表明血糖升高不仅是代谢紊乱的表现,也可能在 AP-ALI/ARDS 中发挥关键作用:高血糖水平可增强氧化应激、促进晚期糖基化终末产物(advanced glycation end products, AGEs)积累,并进一步激发炎症信号通路协同导致屏障功能受损和组织水肿。本研究结果与上述研究一致,提示在 AP 患者血糖水平的监测与控制不仅关乎代谢管理,更是调节免疫炎症反应、减轻组织损伤的潜在策略。重视 AP 严重程度与血糖水平并采取针对性干预措施,有望为延缓或阻断 AP-ALI/ARDS 提供临床契机。

### 3.3 影像组学与深度学习模型的综合性能与临床转化潜力

影像组学特征可以一定程度反映胰腺实质区域的纹理异质性、细节信息以及灰度分布等情况;动脉期特征更敏感地反映胰腺微循环障碍和早期坏死;门静脉期特征则更多反映组织水肿与灌注异常。本研究基于双期影像组学,探讨对 AP-ALI/ARDS 的多时相评估。相较于常规临床特征,影像组学提升模型的预测性能,  $AUC = 0.873$ 。基于样本量,本研究选择精简的 3D ResNet-18 架构构建深度学习模型。这是权衡之后的设计选择,旨在优先保证模型的泛化性能,而非一味追求理论上的特征丰富度。通过特征筛选,保留具有高度判别力的深度特征,深度学习模型  $AUC(0.909)$  优于影像组学模型和临床模型,这提示其可能捕获了除常规纹理特征外更为复杂的图像特征。

本研究进一步通过决策层融合策略,结合临床、影像组学与深度学习三类特征构建 DRC 模型。DRC 模型在训练集中表现出卓越的判别能力 ( $AUC = 0.973$ ),显著超越临床模型的性能 ( $AUC = 0.848$ )。DRC 模型在内部、外部测试集中均展现出优异的泛化能力。校准曲线显示 DRC 模型预测概率与实际结果更为接近,且 DCA 表明在大多数国

值范围内 DRC 模型能提供更高的临床净收益,这些结果共同支持 DRC 模型作为临床辅助决策工具的潜在价值。

### 3.4 研究的局限性及未来展望

本研究仍存在一些局限性:首先,回顾性设计和有限的两中心样本可能引入选择偏倚,未来需要通过前瞻性多中心研究进一步验证。其次,尽管本研究已由经验丰富的医师进行勾画以确保质量,但自动分割技术将是大规模临床应用的前提。第三,模型融合方面采用特征层融合策略,不同融合方法对模型性能的影响有待未来进一步探究。最后,深度学习模型的决策过程缺乏真正的透明度和可解释性,仍是临床应用中需要面对的重要挑战。

### 参考文献

- [1] Forsmark CE, Swaroop VS, Wilcox CM. Acute pancreatitis [J]. *New England Journal of Medicine*, 2016, 375 (20): 1972-1981.
- [2] Luiken I, Eisenmann S, Garbe J, *et al.* Pleuropulmonary pathologies in the early phase of acute pancreatitis correlate with disease severity[J]. *PLoS One*, 2022, 17(2): e0263739.
- [3] Wang Z, Liu J, Li F, *et al.* The gut-lung axis in severe acute Pancreatitis-associated lung injury: The protection by the gut microbiota through short-chain fatty acids[J]. *Pharmacological Research*, 2022, 182: 106321.
- [4] Ge P, Luo Y, Okoye CS, *et al.* Intestinal barrier damage, systemic inflammatory response syndrome, and acute lung injury: a troublesome trio for acute pancreatitis[J]. *Biomedicine & Pharmacotherapy*, 2020, 132: 110770.
- [5] Akbarshahi H, Rosendahl AH, Westergren-Thorsson G, *et al.* Acute lung injury in acute pancreatitis—awaiting the big leap [J]. *Respiratory Medicine*, 2012, 106(9): 1199-1210.
- [6] Verma N, Hochegger B, Mukhopadhyay S, *et al.* Acute lung injury[J]. *Journal of Thoracic Imaging*, 2025, 40(3): e0820.
- [7] Dombrowsky T, Kristensen MØ, Rysgaard S, *et al.* Risk factors for and impact of respiratory failure on mortality in the early phase of acute pancreatitis[J]. *Pancreatology*, 2016, 16 (5): 756-760.
- [8] Fei Y, Gao K, Li WQ. Prediction and evaluation of the severity of acute respiratory distress syndrome following severe acute pancreatitis using an artificial neural network algorithm model [J]. *Hepato-Pancreato-Biliary*, 2019, 21(7): 891-897.
- [9] Jia M, Xu X, Zhou S, *et al.* Prediction of acute lung injury in severe acute pancreatitis by routine clinical data[J]. *European Journal of Gastroenterology & Hepatology*, 2023, 35 (1): 36-44.
- [10] Li L, Liu S, Zhang X, *et al.* Predicting persistent acute respiratory failure in acute pancreatitis: the accuracy of two lung injury indices[J]. *Digestive Diseases and Sciences*, 2023, 68(7): 2878-2889.
- [11] Zhou T, Xie CL, Chen Y, *et al.* Magnetic resonance imaging-based radiomics models to predict early extrapancreatic necro-

- sis in acute pancreatitis [J]. *Pancreas*, 2021, 50 (10): 1368—1375.
- [12] Yu NJ, Li XH, Liu C, *et al.* Radiomics models of contrast-enhanced computed tomography for predicting the activity and prognosis of acute pancreatitis[J]. *Insights into Imaging*, 2024, 15(1):158.
- [13] Wu S, Zhou Q, Cai Y, *et al.* Development and validation of a prediction model for the early occurrence of acute kidney injury in patients with acute pancreatitis[J]. *Renal Failure*, 2023, 45 (1):2194436.
- [14] Liu L, Yu N, Liu T, *et al.* Study on the prediction of liver injury in acute pancreatitis patients by radiomic model based on contrast-enhanced computed tomography[J]. *Current Medical Imaging*, 2024, 20: e15734056307393.
- [15] Banks PA, Bollen TL, Dervenis C, *et al.* Classification of acute pancreatitis—2012: revision of the Atlanta classification and definitions by international consensus[J]. *Gut*, 2013, 62 (1): 102—111.
- [16] Ranieri VM, Rubenfeld GD, Thompson BT, *et al.* Acute respiratory distress syndrome: the Berlin Definition [J]. *JAMA*, 2012, 307(23):2526—2533.
- [17] Alberti P, Pando E, Mata R, *et al.* Evaluation of the modified computed tomography severity index (MCTSI) and computed tomography severity index (CTSI) in predicting severity and clinical outcomes in acute pancreatitis[J]. *Journal of Digestive Diseases*, 2021, 22(1): 41—48.
- [18] Gao W, Yang HX, Ma CE. The value of BISAP score for predicting mortality and severity in acute pancreatitis: a systematic review and meta-analysis [J]. *PLoS One*, 2015, 10 (6): e0130412.
- [19] Whybra P, Zwanenburg A, Andrearczyk V, *et al.* The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights [J]. *Radiology*, 2024, 310(2):e231319.
- [20] Zwanenburg A, Vallières M, Abdalah MA, *et al.* The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping[J]. *Radiology*, 2020, 295(2):328—338.
- [21] Shafiq-Ul-Hassan M, Zhang GG, Latifi K, *et al.* Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels [J]. *Medical Physics*, 2017, 44 (3): 1050—1062.
- [22] Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models[J]. *Medical Decision Making*, 2006, 26(6):565—574.
- [23] Liu J, Zou GJ, Yang L, *et al.* Early prediction of persistent organ failure by circulating endothelial progenitor cells in patients with acute pancreatitis [J]. *Shock*, 2018, 50 (3): 265—272.
- [24] Gao H, Ma X, Wang Z, *et al.* The effects of different types of exercise on pulmonary inflammation and fibrosis in mice with type 2 diabetes mellitus[J]. *Cells*, 2025, 14(13):1026.
- [25] Nobs SP, Kolodziejczyk AA, Adler L, *et al.* Lung dendritic-cell metabolism underlies susceptibility to viral infection in diabetes [J]. *Nature*, 2023, 624(7992):645—652.

(收稿日期:2025—11—10

修回日期:2025—12—28)